

33

CLAIMS

1. A method of generating an index entry for a record in a semi-structured database, the database comprising a plurality of records, each record comprising one or more fields having a plurality of characters therein, the method including the steps of:
- (i) analysing each field in accordance with a predetermined criterion so as to identify an entry within said field;
- (ii) generating at least one index entry representing a concordance between an identified entry and the record corresponding to the identified entry, characterised in that:
- for each of a plurality of predetermined formats, the analysing step (i) further includes:
- searching said field to identify a sequence of characters having a format corresponding to the predetermined format, said identified sequence of characters being deemed to constitute an identified entry.
2. A method according to claim 1, the method further comprising the step of:
- iii) for at least one field, defining any characters not identified as an entry in step (i) as a free text entry.
3. A method according to claim 2, wherein the free text entry comprises at least one free text word defined by a sequence of alphanumeric characters, the method further comprising the steps of:
- iv) identifying at least one free text word in a field by comparing the free text entry with at least one selection criterion defining one or more predetermined characteristics of a free text word; and,
- v) generating a plurality of index entries representing a concordance between the selected free text words determined in step (iv) and the respective records.
4. A method according to ^{claim 1} ~~any one of the preceding claims~~, wherein the records within the semi-structured database are further arranged in groups of records, each group of records being located in a heading field and being identified by at least one

heading entry, wherein the method further comprises the steps of, for each heading field:

vi) identifying heading entries by comparing each heading field with each of a plurality of selection criteria, each selection criterion defining one or more
5 predetermined characteristics of a respective heading entry; and,

vii) generating a plurality of index entries representing a concordance between the heading entries determined in step (vi) and the group of records in the heading field.

10 5. A method according to ~~any one of claims 1 to 4~~, further including the step of arranging the index entries into groups of index entries in accordance with predetermined criteria,

6. Apparatus for generating an index entry for a record in a semi-structured
15 database, the database comprising a plurality of records, each record comprising one or more fields having a plurality of characters therein, the apparatus comprising:

a processor for analysing each field in accordance with a predetermined criterion so as to identify an entry within said field;

an index generator for generating a plurality of index entries representing a
20 concordance between the entries identified by the processor and a record; and

a data store for storing the index entries,
characterised in that

the processor further includes means for searching said field to identify a sequence of characters having a format corresponding to each of a plurality of
25 predetermined formats, said identified sequence of characters being deemed to constitute an identified entry.

SUB BT7 7. Apparatus according to claim 6, wherein for at least one field, the processor
defines any data not determined previously as an entry as a free text entry.

30

8. Apparatus according to claim 7, wherein the free text entry comprises at least one free text word defined by a sequence of alphanumeric characters, wherein the processor identifies at least one selected free text word for a field by comparing

the free text entry with at least one selection criterion defining one or more predetermined characteristics of a selected free text word; and, wherein the index generator generates a plurality of index entries representing a concordance between the selected free text words determined by the processor and the respective records.

5

9. Apparatus according to ~~any one of claims 6 to 8~~, wherein the records within the semi-structured database are further arranged in groups of records, each group of records being located in a heading field and being identified by at least one heading entry, wherein the processor is arranged to identify heading entries by comparing
10 each heading field with each of a number of selection criteria, each selection criterion defining one or more predetermined characteristics of a respective heading entry and wherein the index generator generates a plurality of index entries representing a concordance between the heading entries determined by the processor and the group of records in the heading field, the index entries being stored in the store.

15

10. Apparatus for accessing a semi-structured database in accordance with an input request for information, the semi-structured database having an index generated in accordance with the method of claim 5, the apparatus comprising:

input means for receiving the request;
20 a parser for parsing the request to determine the components of the request;
a slot filler for determining whether the request includes any verb components forming a verb or verb group; and, if the request includes any verb components, the slot filler determines the position of the verb or verb group within the request, and determines any subject components representing the subject of the
25 request and any object components representing the object of the request using the position of the verb or verb group; and, if the request includes no verb components, the slot filler determines any components to be object components, wherein each slot corresponds to one of the group of index entries and wherein the slot filler is arranged to allocate at least one component to a respective slot of a slot-and-filler
30 request; and,
a query constructor for accessing a database,
wherein the query constructor is arranged to compare each of the components allocated to a slot in the slot-and-filler request to one or more index

ADD B87

ADD B9